18th European Symposium on Computer Aided Process Engineering – ESCAPE 18
Bertrand Braunschweig and Xavier Joulia (Editors)

Product data quality in the vaccine industry, a Model-Driven Architecture for interoperability between information systems

Néjib Moalla,[a,b] Abdelaziz Bouras,[a] Yacine Ouzrout,[a] Gilles Neubert[a]

[a]*CERRAL/LIESP. IUT Lumière Lyon 2, 160 Boulevard de l'Université, 69500, Bron, France*
[b]*Sanofi Pasteur / Campus Mérieux. 1541, Avenue Marcel Mérieux 69280, Marcy-L'étoile, France*

Abstract

We present in this paper the benefits of a Model-Driven Architecture (MDA) to ensure the interconnection of different business contexts' specifications by providing a linked structure of models. This enables to generate bridges that connect implementations in different platforms. In this way, the systems interoperability is satisfied.
In fact, the concepts of interoperability are useful to enable the interconnection between different enterprise services and divisions at four levels: business level, process level, service/application level and data level. When the enterprise modeling concepts structure some models to enable the connection between business and process levels, the software engineering concepts structure the models between applications and manipulated data.
The MDA approach is widely considered as a methodology to software generation from models, with a focus on enterprise and business models. Deploying an MDA approach in vaccine industry allows us to deal with business product complexity and regulatory constraints. It helps their translation to some business rules and recommendations when we try to communicate data models from heterogeneous information systems. This will finally ensure product data quality in different information systems throughout the product lifecycle.

Keywords: Vaccine Industry, Model-Driven Architecture, Interoperability, Data Quality.

1. Introduction

The vaccine industry is singular from other pharmaceutical industries by the specificity of its product, the vaccine. Mainly specific by its molecular definition, the product must be manufactured with the closely respect of many strict regulations. Moreover, from a product lifecycle scope, the vaccine product progress with different definitions and formats according to each business contexts in which the product has its local specifications. The vaccine product be so supported by heterogeneous information systems which need to be connected.
We present in the second section an overview of vaccine supply chain specifications. In the third section, we detail our proposed approach for data quality and we discuss in the fourth section achieved results. We present in the last section some conclusions and perspectives for future works.

2. The vaccine supply chain

The specificity of vaccine industry dominates in the particular definition of its product, the vaccine [1]. At the production stage, we can only control the manufacturing process and not the product itself. Indeed, the biological aspect of the active substances in vaccines differs from product in the chemical industry (i.e. pharmaceutical industry) by a very complex structure. The biological production consists of a mixture of molecular substances not always well identified. This specificity makes the control procedures, already different from one product to another, both delicate and complex.

2.1. The vaccine product

Along its lifecycle, a vaccine passes by two huge stages: the vaccine design and the industrial production [2].

2.1.1. Vaccine design

The development of a new vaccine presents a long term action that exceeds rapidly 10 years (figure 1). Starting from exploring phase, researchers look for new substances being able to contribute to creating a vaccine against a given disease. If a vaccine candidate emerges from the discovery research, preclinical tests allow its characterization for a better control of its behavior to generate the appropriate antigens.
In clinical tests stage, several phases allow to better characterize the product by testing an increasing number of patients to determine the effectiveness, the safety and the harmlessness of the vaccine with the suitable intervals of injections. All information concerning clinical reports, tests, and control results are capitalized throughout these phases. To prove the productivity and industrial capability, three consistency batches are manufactured.

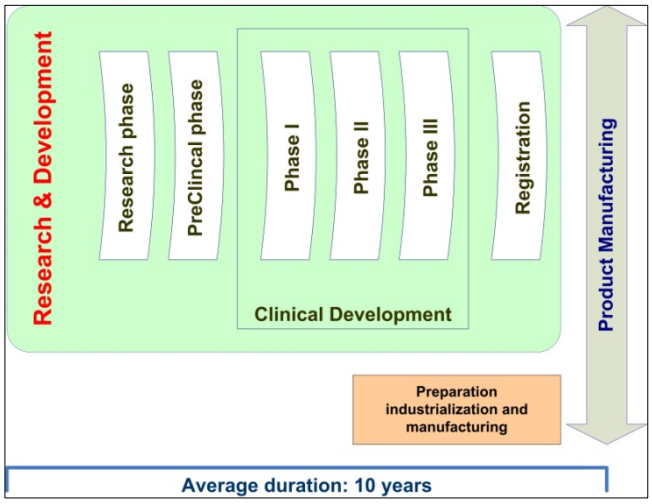

Figure 1: Vaccine design

Finally, to produce and market this new vaccine, it is necessary to prove to the health authorities (WHO http://www.who.int/, FDA in the USA http://www.fda.gov/, etc) the utility of the vaccine [3, 4]. This passes by the submission of Marketing Authorization (MA) Request to the Health authorities of destination countries. The MA contains all the information collected during the process of research and development. Once the product is approved, it can be manufactured and distributed inside this country.

2.1.2. Vaccine production

The vaccine production is a rather complex process lasting from 6 months to 2 years. The production process, performed by fixed batch size, passes by several states. For each state, we identify three tasks, the manufacturing, the quality control and the batch release. We can divide all manufacturing states into two major steps: biological manufacturing and pharmaceutical manufacturing (Figure 2).

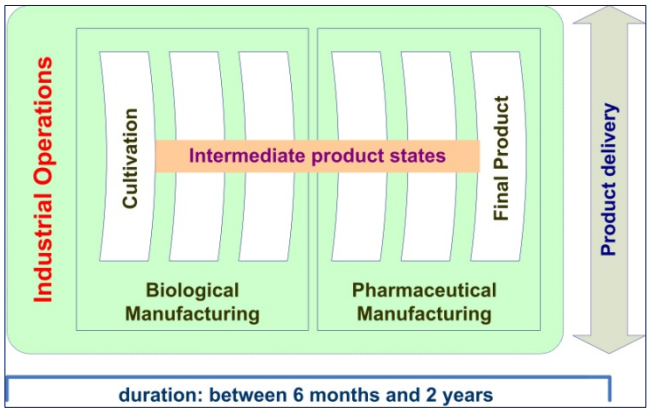

Figure 2: Vaccine manufacturing

- Biological manufacturing: covers from the vaccine valence manufacturing state until the final active substance state (monovalent) using production specificities retained and presented in the MA.
- Pharmaceutical manufacturing: consists of the mixture of active substances, their distribution in the appropriate dosage forms (syringe, blister, etc.) until the packaging. These operations must respect the sterility conditions, quality of air, quality of the final product, etc; all of them are specified in the MA.

2.1.3. Vaccine delivery and product traceability

From a logistic point of view, the vaccine delivery must be done with the respect of customer recommendations specified in the Marketing Authorizations. Delivery lead time must be respected because product shelf life is already carved on the packaging and any modification can breed distribution problems. For product traceability, it is recommended to store batch documents for many years (more than 40 years for blood products). This problem refers to long term data preservation efforts that may be expected in the particular context of vaccine industry.

2.2. The vaccine product data

When we analyze the whole process of vaccine development and production presented previously, we can identify various services ensuring product evolution. Considering the complexity of the biological definition of the vaccine and the large quantity of data to define it, the management of the product data presents a big challenge in this context. In the development stage, product data are generally manually written data, so it's very complex to analyze these data and use them in other information systems. In the production stage, product data definition is distributed between many information systems that interact to produce product according its specifications identified in each product Marketing Authorization. Face to the diversity of business activities and information systems inside, product data quality is rapidly altered when their evolution is not well handled. There is a need for a global and common vision or directive to

follow throughout business contexts to ensure the coherence of product data and maintain their quality inside different information systems.

3. Our approach for data quality

The evolution of processes, systems, services and infrastructures present generally a direct impact on product data. The goal to maintain data quality is always challenged especially in our particular context, highly constrained by regulatory limits and recommendations.

3.1. Data quality

The data quality problem is approached in the literature by the identification of some quality dimension [5-8]. Defined methodologies are more focused on data quality measurement and improvement in information systems solutions than the definition of approaches to handle data quality in networked information systems [9-11]. In this perspective, the concepts of interoperability between information systems allow to deal with business, process, services and applications specificities when we need to ensure data quality. Due to the diversity of information systems architectures and the complexity of process definition in each business context of vaccine industry, we propose to approach the interoperability problem through the proposition of a Model-Driven Architecture (MDA).

3.2. Model-Driven Architecture

The Model-Driven Architecture (MDA) [12] approach marks the new generation of software engineering based on models technologies. In fact, when we need to develop a new software application to be integrated in a specific perimeter, it is possible to take advantages of existing models already optimized for this context to define targeted software application [13]. The MDA approach is identified to be the least expensive and the fastest method to build new applications [14]. Developed by the OMG, the MDA approach consists in defining four abstraction levels [15]:

- The Computational Independent Model (CIM) level is a software independent model used to describe a business system.
- The Platform Independent Model (PIM) level, as a transformation of the CIM, consists of a model with a high level of abstraction that is independent of any implementation technology. Within a PIM, the system is modeled from the viewpoint of how it best supports the business.
- The Platform Specific Model (PSM) level, as a transformation of the PIM, is tailored to specify the system in terms of the implementation constructs that are available in one specific implementation technology.
- The final step in the development is the transformation of each PSM into code. Because a PSM fits its technology rather closely, this transformation is relatively straightforward.

From one abstraction level to another, the MDA Framework structures the transformation process and ensures the mapping of metamodels expressed in MOF as the metametamodel language defined by the OMG [16].

3.3. MDA for data quality

To ensure data quality on a given software application, we need to identify and structure the impact of all entities that may affect data quality throughout product lifecycle. Our approach aims in its perspective to develop a new software application to manage product data lifecycle. The deployment of the MDA Framework allows to integrate all business models, already defined and optimized in different business contexts, in one

computer independent model. Through the transformation process, the generated Platform Specific Model is intended to be transformed to obtain the code of a new application. However, we propose in our particular context to make use of this PSM expressed in the same platform than the information system where we need to ensure data quality. Data quality is achieved when we compare the new model with the exiting one. In the mapping process, we aim to feed existent application models with all entities needed to maintain data quality under a global vision of product data lifecycle. We present in figure 3 an overview of our architecture when applied to ensure data quality in the Enterprise Resource Planning (ERP) system.

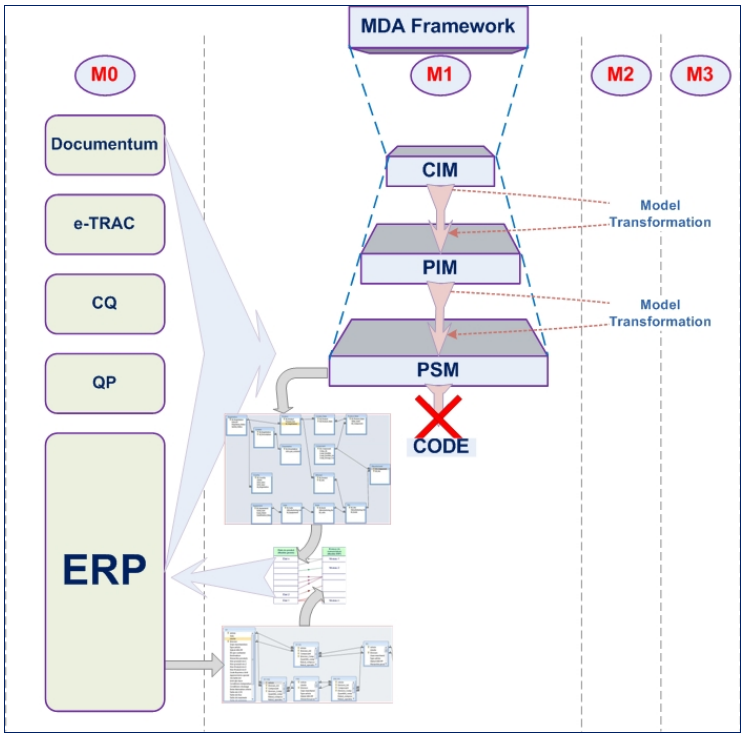

Figure 3: Proposed architecture for data quality

4. Discussion

The application of MDA concept to ensure interoperability between information systems presents a new research orientation. Through the definition of this new context, our approach to maintain data quality needs some specific efforts to handle the models. In fact, at each MDA abstraction level, the generated models must be reviewed and optimized to verify that they reflect the reached objectives. Moreover, the model transformation process faces some problems in the mapping process between metamodels when expressed in specific languages relatively to transformation platforms (KM3 for ATL and KMT for KerMeta within ECLIPSE).

The deployment of the approach allows to ensure the quality of some critical data in the ERP system. Indeed, the recommendation issued from mapping process between the data model of the ERP and the generated one at PSM level can suggest some evolutions of the ERP model and not just its data. In this case, some alternative solutions must be

found to maintain the stability of the ERP system and its conformity to regulatory recommendations.

5. Conclusion

In this paper, we highlight the vaccine supply chain context with a specific focus on its product, the vaccine. Due to the complexity of the vaccine definition and the constraints imposed by health authorities, data quality is considered as a big challenge to deal with. When product data are part of various information systems, the concept of interoperability is approached via the definition of a Model-Driven Architecture to deal with different existent models and generate a new Platform Specific Model to be used to validate data quality in a specific perimeter. With the forthcoming steps, we aim to extend proposed models to include more specific product details to structure different system evolutions to be projected directly on product data.

Acknowledgment

This work is done within the collaboration with Sanofi Pasteur France company where Néjib MOALLA is integrated as a researcher in Industrial Operations services during his PhD thesis (CIFRE N° 865/2005).

References

[1] FDA, "Challenge and Opportunity on the Critical Path to New Medical Products," Food And Drung Administration March 2004 (revised).
[2] E. Salinsky and C. Werble, "The Vaccine Industry: Does It Need a Shot in the Arm?," The George Washington University, Background Paper January 25, 2006.
[3] C. Grace, "Global health partnership impact on commodity pricing and security," DFID Health Resource 2004.
[4] S. Thaul, "Vaccine policy issues for the 108h congress," The library of Congress updated May 22 2003.
[5] A. R. Tupek, "Definition of Data Quality," Census Bureau 2006.
[6] L. L. Pipino, Y. W. Lee, and R. Y. Yang, "Data Quality Assessment," *COMMUNICATIONS OF THE ACM,* vol. 45, p. 8, 2002.
[7] A. D. Chapman, "PRINCIPLES OF DATA QUALITY," Global Biodiversity Information Facility, Copenhagen 2005.
[8] M. Martin, "Measuring and Improving Data Quality. Part 3: Improving Data Quality," NAHSS June 2005.
[9] C. Batini and M. Scannapieca, *Data Quality: Concepts, Methodologies and Techniques*: Springer-Verlag Berlin Heidelberg, 2006.
[10] R. Y. Wang, M. Ziad, and Y. W. Lee, *DATA QUALITY*: Kluwer Academic Publishers, 2002.
[11] J. E. Olsen, *Data Quality: The Accuracy Dimension*: Morgan Kaufmann Publishers, 2003.
[12] J. Miller and J. Mukerji, "MDA Guide," OMG 12th June 2003.
[13] H. Kadima, *MDA : Conception orientée objet guidée par les modèles*, DUNOD ed.: DUNOD, 2005.
[14] O. Pastor and J. C. Molina, *Model-Driven Architecture in Practice: A Software Production Environment Based on Conceptual Modeling*, First Edition ed.: Springer, 2007.
[15] X. Blane, *MDA en action*, Eyrolles ed.: Eyrolles, 2005.
[16] A. Kleppe, J. Warmer, and W. Bast, *MDA Explained: The Model Driven Architecture™: Practice and Promise* vol. 1: Addison Wesley, 2003.